Quality Descriptions

DDI addresses descriptions of quality at three levels, the quality of the metadata as captured, means of ensuring policy for processing, and the quality of the data. Statements regarding the quality standards used for processing or measuring the quality of the data related to the metadata may be referenced from the following areas: study unit, group, sub-group, resource package, methodology, various descriptions of data processing, data file creation or copying, and archival processing.

# Metadata Quality

Metadata quality is captured at the level of the Maintainable object and refers to the metadata contained in that object. It is held in the complex element MetadataQuality which is found in AbstractMaintainable. It is a relatively simple statement of quality such as completeness, transcription of content from the original source, level of review or verification, etc.

MetadataQuality

TypeOfMetadataQuality (1..1)

MeasurePurpose (0..1)

MeasureValue (0..1)

Description (0..1)

Both TypeOfMetadataQuality and MeasureValue are CodeValueType and support the use of an external controlled vocabulary. These are the two primary pieces of information needed to track metadata quality internally within an organization. TypeOfMetadataQuality identifies the measure being tracked and MeasureValue provides a specific value for that measure. For example, an organization may have a number of measures one of which rates its transcription quality. It may have various values for this including: 1) Tagging and transfer of a digital content, 2) double entry, 3) single entry direct transcription, 4) selected transcription, and 5) summarization. The purpose for capturing the measure is held in MeasurePurpose (StructuredStringType) and other descriptive information that is useful to the organization or in particular a user, is held in Description (StructuredStringType).

EXAMPLE:

<r:MetadataQuality>

<r:TypeOfMetadataQuality codeListName=”MPCQualityTracking”>

Transcription

</r:TypeOfMetadataQuality>

<r:MeasurePurpose>

<r:Content>To identify the transcription method used for the metadata in this section</r:Content>

</r:MeasurePurpose>

<r:MeasureValue codeListName=”MPCTranscriptionType”>1</r:MeasureValue>

<r:Description>

<r:Content>Content was original digital text. Content was tagged and processed through an XSLT to transform it to DDI structure.</r:Content>

</r:Description>

</r:MetadataQuality>

# Quality Statement

Quality statements are compiled in a QualityStatementScheme which may be published within a StudyUnit, Group, or ResourcePackage. Quality statements primarily address processes and steps that are taken to ensure quality within those processes. A QualityStatement allows for either the identification of an external standard plus a statement regarding compliance with that standard, or a general statement of steps taken to ensure quality for a given process or activity. A QualityStatement is attached to a processing area by reference from the description of the process/activity itself.

QualityStatement

Extension base: Versionable

CHOICE: (0..n)

StandardsCompliance

Standard (1..1)

ComplianceDescription (0..1)

OtherQualityStatement

ENDCHOICE

StandardsCompliance consists of a reference to the Standard using the structure of OtherMaterial. This could reference a document, web site, or other source containing a formal standard for processing, best practice, internal protocol, or other statement of quality. ComplianceDescription (StructuredStringType) provides details on how this standard or protocol was applied, in particular noting any deviations or issues that would have an impact on the quality factors being assessed. When no formal standard or protocol exists, use OtherQualityStatement (StructuredStringType) to describe steps taken to ensure quality. Quality statements can be referenced from the following locations and should relate quality assessment information focused on the process, activity or general coverage area where the reference is included:

|  |  |
| --- | --- |
| a:ArchiveSpecific/r:QualityStatementReference | In this location the QualityStatement being referenced should relate to the monitoring of the Archive's activities and operations. For example: TRAC certification using the reference to the TRAC standard and noting the Archive's certification dates in the compliance section |
| d:Methodology/r:QualityStatementReference | In this location the QualityStatement being referenced should relate to overall methodology, i.e., overall scientific method followed, statistical standards, |
| d:CollectionEvent/r:QualityStatementReference | In this location the QualityStatement being referenced should relate to a specific collection event, i.e., respondent confidentiality protocol, use of human subjects protocol |
| d:ProcessingEvent/r:QualityStatementReference | In this location the QualityStatement being referenced should relate to the specific processing event, i.e., best practices for data cleaning in a specific area |
| g:ResourcePackage/r:QualityStatementReference | In this location the QualityStatement being referenced should relate to the specific activities surrounding the creation of the resource package, i.e., criteria for selecting the schemes to include, review processes, modification, etc. |
| g:ResourcePackage/r:QualityStatementScheme | The inline content of a publication package not related to a specific study (payload) |
| g:ResourcePackage  /r:QualityStatementSchemeReference | The referenced content of a publication package not related to a specific study (payload) |
| g:Group/r:QualityStatementReference | In this location the QualityStatement being referenced should relate to the specific activities surrounding the creation of the Group, i.e., grouping, selection, or organizational policies. |
| g:Group/r:QualityStatementScheme | The scheme of QualityStatements shared by the Group included in-line |
| g:Group/r:QualityStatementSchemeReference | The scheme of QualityStatements shared by the Group included by reference |
| g:SubGroup/r:QualityStatementReference | In this location the QualityStatement being referenced should relate to the specific activities surrounding the creation of the SubGroup, i.e., grouping, selection, or organizational policies. (Allowing for the later subordination of a Group thereby turning it into a SubGroup) |
| g:SubGroup/r:QualityStatementSchemeReference | The scheme of QualityStatements shared by the sub-group included by reference |
| pi:PhysicalInstance/r:QualityStatementReference | In this location the QualityStatement being referenced should relate to the data file, i.e., a statement regarding data lost during rescue from an old media, the use of standards protocols in verification that this is a valid copy of another file |
| s:StudyUnit/r:QualityStatementReference | In this location the QualityStatement being referenced should relate to the specific activities surrounding the creation of the StudyUnit, i.e., coverage, purpose, organization, etc. |
| s:StudyUnit/r:QualityStatementScheme | The scheme of QualityStatements shared by the StudyUnit included in-line |
| s:StudyUnit/r:QualityStatementSchemeReference | The scheme of QualityStatements shared by the StudyUnit included by reference |

EXAMPLES:

<r:QualityStatementScheme isMaintainable=”true” typeOfIdentifier=”Canonical” scopeOfUniqueness=”Maintainable”>

<r:URN>urn:ddi:us.archive:QScheme\_1:1</r:URN>

<r:QualityStatement isVersionable=”true” ” typeOfIdentifier=”Canonical” scopeOfUniqueness=”Maintainable”>

<r:URN>urn:ddi:us.archive:QScheme\_1.QS\_1:1</r:URN>

<r:StandardsCompliance>

<r:Standard isIdentifiable=”true” typeOfIdentifier=”Canonical” scopeOfUniqueness=”Maintainable”>

<r:URN>urn:ddi:us.archive:QScheme\_1.Standard\_a:1</r:URN>

<r:TypeOfMaterial>Standard</r:TypeOfMaterial>

<r:Citation>

<r:Title>Reference Model for an Open Archival Information System (OAIS): Recommended Practice CCSDS 650.0-M-2</r:Title>

</r:Citation>

<r:ExternalURNReference> http://public.ccsds.org/publications/archive/650x0m2.pdf

</r:ExternalURNReferenc>

</r:Standard>

<r:ComplianceDescription>

<r:Content>The archive complies with the OAIS model regarding the tracking of information from SIP through data processing and archival management to the creation of the AIP and assembly of the DIP.</r:Content>

</r:ComplianceDescription>

</r:StandardsCompliance>

</r:QualityStatement>

<r:QualityStatement isVersionable=”true” ” typeOfIdentifier=”Canonical” scopeOfUniqueness=”Maintainable”>

<r:URN>urn:ddi:us.archive:QScheme\_1.QS\_2:1</r:URN>

<r:StandardsCompliance>

<r:Standard isIdentifiable=”true” typeOfIdentifier=”Canonical” scopeOfUniqueness=”Maintainable”>

<r:URN>urn:ddi:us.archive:QScheme\_1.Standard\_b:1</r:URN>

<r:TypeOfMaterial>Policy</r:TypeOfMaterial>

<r:Citation>

<r:Title>Univerity Of Minnesota Board Of Regents Policy: Reserch Involving Human Subjects</r:Title>

</r:Citation>

<r:ExternalURNReference>

http://www1.umn.edu/regents/policies/academic/HumanSubjects.pdf

</r:ExternalURNReferenc>

</r:Standard>

<r:ComplianceDescription>

<r:Content>Policy strictly complied to.</r:Content>

</r:ComplianceDescription>

</r:StandardsCompliance>

</r:QualityStatement>

<r:QualityStatement isVersionable=”true” ” typeOfIdentifier=”Canonical” scopeOfUniqueness=”Maintainable”>

<r:URN>urn:ddi:us.archive:QScheme\_1.QS\_3:1</r:URN>

<r:OtherQualityStatement>

<r:Content>As data is ported from one format to another, the following checks are made to verify the accurate transfer of the data items. Record Count and frequency count on all data items at the category level. A random sample of records is pulled from each file for item by item comparison. </r:Content>

</r: OtherQualityStatement >

</r:QualityStatement>

</r:QualityStatementScheme>

<a:Archive isMaintainable=”true” typeOfIdentifier=”Canonical” scopeOfUniqueness=”Maintainable”>

<r:URN>urn:ddi:us.archive:Archive\_1:1</r:URN>

<a:ArchiveSpecific>

<r:QualityStatementReference isReference=”true” isExternal="true" lateBound="false" typeOfIdentifier="Canonical">

<r:URN>urn:ddi:us.archive:QScheme\_1.QS\_1:1</r:URN>

<r:TypeOfObject>QualityStatement</r:TypeOfObject>

</r:QualityStatementReference>

<a:ArchiveSpecific>

</a:Archive>

<d:DataCollection isMaintainable=”true” typeOfIdentifier=”Canonical” scopeOfUniqueness=”Maintainable”>

<r:URN>urn:ddi:us.archive:DataColl\_1:1</r:URN>

<d:CollectionEvent isIdentifiable=”true” typeOfIdentifier=”Canonical” scopeOfUniqueness=”Maintainable”>

<r:URN>urn:ddi:us.archive:DataColl\_1.CE\_1:1</r:URN>

<r:QualityStatementReference isReference=”true” isExternal="true" lateBound="false" typeOfIdentifier="Canonical">

<r:URN>urn:ddi:us.archive:QScheme\_1.QS\_2:1</r:URN>

<r:TypeOfObject>QualityStatement</r:TypeOfObject>

</r:QualityStatementReference>

</d:CollectionEvent>

</d:DataCollection>

<pi:PhysicalInstance isMaintainable=”true” typeOfIdentifier=”Canonical” scopeOfUniqueness=”Maintainable”>

<r:URN>urn:ddi:us.archive:PhysInst\_1:1</r:URN>

<r:QualityStatementReference isReference=”true” isExternal="true" lateBound="false" typeOfIdentifier="Canonical">

<r:URN>urn:ddi:us.archive:QScheme\_1.QS\_3:1</r:URN>

<r:TypeOfObject>QualityStatement</r:TypeOfObject>

</r:QualityStatementReference>

</pi:PhysicalInstance>

# Data Quality

The quality of the data collected is addressed in the ProcessingEvent containing DataAppraisalInformation. This contains two common measures of data quality from survey methodology, ResponseRate and SamplingError. It also allows for the description of other quality measures in OtherAppraisalProcess. Note that ProcessingEvent also contains a QualityStatementReference. This should be used to relay information on process quality. DataAppraisalInformation is used primarily to capture the results of data appraisal measures.

DataAppraisal

ResponseRate (0..n)

SampleSize (0..1)

NumberOfResponses (0..1)

SpecificResponseRate (0..1)

r:Description (0..1)

SamplingError (0..n)

OtherAppraisalProcess (0..n)

The response rate can be repeated to express differing response rates by mode of deliver, location, etc. The individual response rates may be expressed as a total sample size with number of responses and/or as a specific response rate. The description should be used to differentiate when multiple response rates are provided. SamplingError (StructuredStringType) is intended to contain a discussion of the sampling error. It may be structured to differentiate between the statement of the error itself, how it was calculated, etc. OtherAppraisalProcess allows for the description of other measures of data appraisal as needed.

Example:

<d:DataAppraisal>

<d:ResponseRate>

<d:SampleSize>5000</d:SampleSize>

<d:NumberOfResponses>3768</d:NumberOfResponses>

<d:SpecificResponseRate>.7536</d:SpecificResponseRate

</d:ResponseRate>

<d:SamplingError><r:Content><xhtml:b>Calculation of Standard Errors <xhtml:br/>Totals and percentages. </xhtml:b>Tables A through C in this chapter contain the necessary information for calculating the standard errors of sample estimates in this data product. To calculate the standard error, it is necessary to know:<xhtml:ul><xhtml:li>The unadjusted standard error for the characteristic (given in Table A for estimated totals or Table B for estimated percentages) that would result under a simple random sample design of people, housing units, households, or families.</xhtml:li><xhtml:li>The design factor for the particular characteristic estimated (given in Table C) based on the sample design and estimation techniques employed to produce long form data estimates. </xhtml:li><xhtml:li>The number of people, housing units, households, or families in the publication area. </xhtml:li><xhtml:li>The observed sampling rate.</xhtml:li></xhtml:ul>

The design factor is the ratio of the estimated standard error to the standard error of a simple random sample. The design factors reflect the effects of the actual sample design and the complex ratio estimation procedure used for Census 2000….</r:Content><d:SamplingError>

</d:DataAppraisal>